

Bundesgesundheitsbl 2024 · 67:710–720
<https://doi.org/10.1007/s00103-024-03884-8>
 Eingegangen: 14. Dezember 2023
 Angenommen: 19. April 2024
 Online publiziert: 15. Mai 2024
 © The Author(s) 2024



Dagmar Waltemath¹ · Oya Beyan² · Katrin Cramer³ · Angela Dedié⁴ · Kerstin Gierend⁵ · Petra Gröber⁶ · Esther Thea Inau¹ · Lea Michaelis¹ · Ines Reinecke⁷ · Martin Sedlmayr⁸ · Sylvia Thun⁹ · Dagmar Krefting¹⁰

¹ Abteilung Medizininformatik, Institut für Community Medicine, Greifswald, Deutschland; ² Medizinische Fakultät und Uniklinik Köln, Institut für Biomedizinische Informatik, Universität zu Köln, Köln, Deutschland; ³ Schweizerisches Institut für Bioinformatik, Personalisierte Gesundheitsinformatik, Basel, Schweiz; ⁴ Deutsches Zentrum für Diabetesforschung (DZD), Geschäftsstelle am Helmholtz Zentrum München, München, Deutschland; ⁵ Abteilung für Biomedizinische Informatik am Zentrum für Präventivmedizin und Digitale Gesundheit (CPD), Medizinische Fakultät Mannheim der Universität Heidelberg, Mannheim, Deutschland; ⁶ Datenintegrationszentrum Universitätsmedizin Rostock, Rostock, Deutschland; ⁷ Datenintegrationszentrum, Zentrum für Medizinische Informatik, Universitätsklinikum Carl Gustav Carus Dresden, Dresden, Deutschland; ⁸ Institut für Medizinische Informatik und Biometrie, Med. Fakultät Carl Gustav Carus, TU Dresden, Dresden, Deutschland; ⁹ Berliner Institut für Gesundheitsforschung in der Charité – Universitätsmedizin Berlin, Berlin, Deutschland; ¹⁰ Institut für Medizinische Informatik, Universitätsmedizin Göttingen und Deutsches Zentrum für Herz-Kreislauf-Forschung, Partner Site Göttingen, Göttingen, Deutschland

FAIRe Gesundheitsdaten im nationalen und internationalen Datenraum

Hintergrund

Gesundheitsdaten haben in der heutigen datenorientierten Welt einen hohen Stellenwert. Durch automatisierte Verarbeitung können beispielsweise Prozesse im Gesundheitswesen optimiert und klinische Entscheidungen unterstützt werden. Dabei ist es wichtig, dass die Daten aussagekräftig, von hoher Qualität und vertrauenswürdig sind. Nur so kann eine sinnvolle automatische Verarbeitung, aber auch Nachnutzung garantiert werden. Die Erhebung von Primärdaten – sei es im Behandlungs- oder Studienkontext – ist ein aufwendiger, kosten- und zeitintensiver Prozess, der verteilt über verschiedene Anwendungssysteme und auch Klinik- und Landesgrenzen hinweg stattfindet. Sekundärdatennutzung wiederum macht moderne medizinische Forschung, die auf große Datensätze angewiesen ist, erst möglich, beispielsweise für das Trainieren künstlicher Intelligenz (KI).

Der Datenbereitstellung und -nachnutzung widmen sich verschiedene Initiativen in Deutschland. Neben der Medizininformatik-Initiative (MII; [1]), welche fachübergreifend strukturierte

Daten aus den Krankenhausinformationssystemen in den lokalen Datenintegrationszentren der Universitätskliniken verfügbar macht, gibt es zahlreiche fachspezifische Initiativen. Beispiele sind das Radiologienetzwerk RACOON (Radiological COOperative Network; [2]), die COMPASS („coordination on mobile pandemic apps best practice and solution sharing“; [3])-Plattform¹ für den Einsatz von Gesundheits-Apps in Pandemiesituationen, das Notaufnahmeregister AKTIN (Aktionsbündnis zur Verbesserung der Kommunikations- und Informationstechnologie in der Intensiv- und Notfallmedizin; [4]) im Rahmen des Netzwerks Universitätsmedizin (NUM; [5]) oder RADARplus (Routine Anonymized Data for Advanced Health Services Research; [6]) für die Nutzung von Daten aus Hausarztpraxen. Das Deutsche Zentrum für Herz- und Kreislauferkrankungen (DZHK) baut für die Sekundärnutzung von Studien- daten seit 2012 die sogenannte Heart Bank auf, deren Plattform seit 2020 auch für die nationalen COVID-19-Kohorten

genutzt und nun als Infrastruktur im NUM weiterentwickelt wird [7, 8].

Die gemeinsame Nutzung von Gesundheitsdaten für die Forschung gelingt dann, wenn die Daten und die beteiligten datenhaltenden Systeme interoperabel sind, d. h. von verschiedenen IT-Systemen und Datenbanken verarbeitet und interpretiert werden können. Dabei ist es wichtig, dass auch Bearbeitungsprozesse und die Herkunft der Daten dokumentiert werden. Dies wird als Provenienz bezeichnet. Provenienz schafft dadurch einen Mehrwert, indem sie Wiederverwendungsmöglichkeiten, aber auch Grenzen aufzeigt und Vertrauen in die Daten schafft. Dabei ist die gleichzeitige Einhaltung der gesetzlichen Datenschutzanforderungen essenziell und muss durch geeignete Maßnahmen, wie z. B. eine zuverlässige Deidentifizierung, sichergestellt werden [9, 10].

Konkrete Anforderungen an die Beschreibung und Kodierung von nachnutzbaren Daten werden in den sogenannten FAIR-Prinzipien beschrieben (FAIR steht als Akronym für Findable = auffindbar, Accessible = erreichbar, Interoperable = interoperabel, Reusable = nachnutzbar; [11]). Diese Anforderun-

¹ <https://num-compass.science/>, Zugegriffen: 26. Februar 2024.

gen entlang des Datenlebenszyklus sind in den vergangenen Jahren zu einem internationalen Maßstab für wiederverwendbare Daten geworden. Eine Positionierung zu den FAIR-Kriterien wird bei Drittmittelanträgen in der Regel als Teil des Forschungsdatenmanagements (FDM) gefordert oder ist explizit Teil von Datenmanagementplänen. Die Europäische Kommission erwartet darüber hinaus die Stellungnahme zur Umsetzung dieser Maßnahmen in den Zwischenberichten. Hierdurch soll das Bewusstsein für die nachvollziehbare Bereitstellung von wissenschaftlichen Ergebnissen gestärkt und die Reproduzierbarkeit von Analyseergebnissen verbessert werden. Anstoß für die sehr dynamische Entwicklung im Gebiet des FDM waren sicherlich die sogenannte Reproduzierbarkeitskrise² und die Digitalisierungsinitiativen des Bundes. In deren Folge sind in den Bundesländern Netzwerke für FDM entstanden, die die Aufgaben auf Landesebene organisieren und bundesweite Strategien mitentwickeln [12].

Verschiedene nationale Forschungsverbände und Infrastrukturprojekte im Gesundheitswesen haben sich bereits klar zu den FAIR-Prinzipien positioniert: Sowohl die Infrastrukturen der MII als auch die der NUM operieren explizit auf Basis der FAIR-Prinzipien. Die MII setzt mit verschiedenen Maßnahmen die FAIR-Kriterien um. Die Auffindbarkeit wird beispielsweise durch das zentrale Deutsche Forschungsdatenportal für Gesundheit³ (FDPG) gewährleistet. Erreichbarkeit und Interoperabilität werden unter anderem durch den gemeinsamen modularen Kerndatensatz (KDS) erreicht, welcher die klinischen Daten aus den Primärsystemen über sogenannte ETL(extrahieren, transformieren, laden)-Prozesse in ein interoperables Format überführt und

für die Sekundärnutzung bereitstellt [1]. Dabei legt der verwendete Datenaustauschstandard HL7 FHIR (Health Level 7 Fast Healthcare Interoperability Resources) nicht nur das Datenformat fest, was als syntaktische Interoperabilität bezeichnet wird, sondern definiert mithilfe von Terminologien wie der Systematisierten Nomenklatur der Medizin (SNOMED), der Logical Observation Identifiers Names and Codes (LOINC) und der International Classification of Diagnoses (ICD) auch die Bedeutung der Datenelemente und einzelnen Informationen. Dies wird als semantische Interoperabilität bezeichnet. Für die Nachnutzbarkeit wurden eine harmonisierte Patienteneinwilligung sowie vertragliche Rahmenbedingungen im sogenannten Teilnahmevertrag formuliert. Die Nationale Forschungsdateninfrastruktur für personenbezogene Gesundheitsdaten (NFDI4Health; [13]) entwickelt neben einem Metadatenschema (MDS) für epidemiologische und Gesundheitsstudien [14] mit den entsprechenden FHIR auch einen zentralen Study-Hub⁴ [15] und unterstützt somit die Auffindbarkeit und Exploration von Studieninformationen. Die FAIRe Bereitstellung von Daten sowie deren Nachnutzung waren 2 Kriterien bei der Erstellung des Basisdatensatzes für klinische Parameter in der Diabetes- und Stoffwechselforschung des Deutschen Zentrums für Diabetesforschung (DZD; [16]).

Im Idealfall können die aus den Daten gewonnenen Erkenntnisse wieder in die Versorgung zurückgespielt werden und dort zur verbesserten Behandlung von Patientinnen und Patienten beitragen [1]. Solche Feedbackmechanismen in die Klinik führen zunehmend zu einer engen Verzahnung von Versorgungs- und Forschungs-IT. Als Beispiel sei hier eine IT-Referenzarchitektur für Datenintegrationszentren aus dem SMITH(Smart Medical Technology for Healthcare; [17])-Konsortium genannt. Teil dieser Architektur ist eine klinische Domäne, die sämtliche Daten aus den klinischen Primärsystemen entgegennimmt und diese in den interoperablen, von der MII defi-

nierten KDS abbildet. Die klinische Domäne enthält also ein interoperables Abbild der Behandlungsdaten, mit dem z. B. klinische Entscheidungsunterstützungssysteme entwickelt und dann bei Erfolg direkt in der Versorgung verwendet werden können.

Nicht direkt in den FAIR-Kriterien adressiert sind die Aspekte Datenqualität und Reproduzierbarkeit. Wichtige Indikatoren für Datenqualität sind Vollständigkeit, Korrektheit und Konsistenz. Speziell für die Qualitätsprüfung der Forschungsdaten aus der klinischen Routine wurden verschiedene Systeme und Indikatoren entwickelt, die mehr oder weniger eng an den klinischen Routinesystemen gemessen werden [18, 19]. Hier ist ein Vorteil von HL7 FHIR, dass die beschriebenen Metadaten direkt in FHIR integriert werden können, wodurch in der Regel eine höhere Qualität der Daten erreicht wird. Reproduzierbarkeit erscheint in digitalen Anwendungen im Vergleich zu z. B. Laborexperimenten relativ einfach: Die gleichen Eingangsdaten sollen mit den gleichen Analysemethoden die gleichen Ergebnisse liefern. In der Realität können unterschiedliche Softwareversionen und unterschiedliche Hardware zu unterschiedlichen Ergebnissen führen. Reproduzierbarkeit erfordert deshalb neben der oben genannten Provenienz auch eine umfangreiche Dokumentation und idealerweise eine Archivierung der in der Analyse genutzten Software und Ausführungsumgebungen. Grundsätzlich kann dies durch die Anwendung der FAIR-Kriterien auf alle am Datenlebenszyklus beteiligten digitalen Objekte erreicht werden (■ Abb. 1).

Wege zu FAIRen Forschungsergebnissen

Die FAIR Guiding Principles for Data Stewardship [11] wurden 2016 publiziert. Für auffindbare, erreichbare, interoperable und nachnutzbare Forschungsdaten werden in einem Kriterienkatalog verschiedene Anforderungen an die Daten spezifiziert (■ Tab. 1; [20–23]). Beispiele für die Umsetzung dieser Kriterien sind die genaue Beschreibung der Daten mit Metadaten, die Nutzung eines global einzigartigen, beständigen Identifier für die

² Die sogenannte Reproduzierbarkeitskrise (Replikationskrise, Krise der Reproduzierbarkeit) bezeichnet die seit 2011 mehrfach systematisch belegte Beobachtung, dass publizierte Erkenntnisse durch die Wiederholung der beschriebenen Studien und Experimente nicht nachvollzogen werden können.

³ <https://forschen-fuer-gesundheit.de/>, Zugegriffen: 26. Februar 2024.

⁴ <https://csh.nfdi4health.de>, Zugegriffen: 26. Februar 2024.

Bundesgesundheitsbl 2024 · 67:710–720 <https://doi.org/10.1007/s00103-024-03884-8>
© The Author(s) 2024

D. Waltemath · O. Beyan · K. Cramer · A. Dedié · K. Gierend · P. Gröber · E. T. Inau · L. Michaelis · I. Reinecke · M. Sedlmayr · S. Thun · D. Krefting

FAIRe Gesundheitsdaten im nationalen und internationalen Datenraum

Zusammenfassung

Gesundheitsdaten haben in der heutigen datenorientierten Welt einen hohen Stellenwert. Durch automatisierte Verarbeitung können z. B. Prozesse im Gesundheitswesen optimiert und klinische Entscheidungen unterstützt werden. Dabei sind Aussagekraft, Qualität und Vertrauenswürdigkeit der Daten wichtig. Nur so kann garantiert werden, dass die Daten sinnvoll nachgenutzt werden können. Konkrete Anforderungen an die Beschreibung und Kodierung von Daten werden in den FAIR-Prinzipien beschrieben. Verschiedene nationale Forschungsverbände und Infrastrukturprojekte im Gesundheitswesen haben sich bereits klar zu den FAIR-Prinzipien positioniert: Sowohl die Infrastrukturen der Medizininformatik-Initiative als auch des

Netzwerks Universitätsmedizin operieren explizit auf Basis der FAIR-Prinzipien, ebenso die Nationale Forschungsdateninfrastruktur für personenbezogene Gesundheitsdaten oder das Deutsche Zentrum für Diabetesforschung. Um eine FAIRe Ressource bereitzustellen, sollte zuerst in einem Assessment der FAIRness-Grad festgestellt werden und danach die Priorisierung für Verbesserungsschritte erfolgen (FAIRification). Seit 2016 wurden zahlreiche Werkzeuge und Richtlinien für beide Schritte entwickelt, basierend auf den unterschiedlichen, domänenspezifischen Interpretationen der FAIR-Prinzipien. Auch die europäischen Nachbarländer haben in die Entwicklung eines nationalen Rahmens für semantische Interoperabilität

im Kontext der FAIR-Prinzipien investiert. So wurden Konzepte für eine umfassende Datenanreicherung entwickelt, um die Datenanalyse beispielsweise im Europäischen Gesundheitsdatenraum oder über das Netzwerk der Observational Health Data Sciences and Informatics zu vereinfachen. In Kooperation mit internationalen Projekten, wie z. B. der European Open Science Cloud, wurden strukturierte FAIRification-Maßnahmen für Gesundheitsdatensätze entwickelt.

Schlüsselwörter

Datenmanagement · Interoperabilität · Datenintegration · IT-Infrastruktur · Netzwerk Universitätsmedizin

FAIR health data in the national and international data space

Abstract

Health data are extremely important in today's data-driven world. Through automation, healthcare processes can be optimized, and clinical decisions can be supported. For any reuse of data, the quality, validity, and trustworthiness of data are essential, and it is the only way to guarantee that data can be reused sensibly. Specific requirements for the description and coding of reusable data are defined in the FAIR guiding principles for data stewardship. Various national research associations and infrastructure projects in the German healthcare sector have already clearly positioned themselves on the FAIR principles: both the infrastructures of the Medical Informatics Initiative and the University Medicine Network operate explicitly on the

basis of the FAIR principles, as do the National Research Data Infrastructure for Personal Health Data and the German Center for Diabetes Research.

To ensure that a resource complies with the FAIR principles, the degree of FAIRness should first be determined (so-called FAIR assessment), followed by the prioritization for improvement steps (so-called FAIRification). Since 2016, a set of tools and guidelines have been developed for both steps, based on the different, domain-specific interpretations of the FAIR principles.

Neighboring European countries have also invested in the development of a national framework for semantic interoperability in the context of the FAIR (Findable, Accessible,

Interoperable, Reusable) principles. Concepts for comprehensive data enrichment were developed to simplify data analysis, for example, in the European Health Data Space or via the Observational Health Data Sciences and Informatics network. With the support of the European Open Science Cloud, among others, structured FAIRification measures have already been taken for German health datasets.

Keywords

Data management · Interoperability · Data integration · IT infrastructure · German Network University Medicine

Auffindbarkeit (F1), die Nutzung von offenen Kommunikationsstandards für die Erreichbarkeit (A1), die Verwendung von FAIRen Vokabularien (I2), formalen, zugänglichen, gemeinsamen und allgemein anwendbaren Wissenssprachen und qualifizierten Referenzen zu anderen Daten für die Interoperabilität (I3) sowie die Angabe einer Nutzungslizenz (R1.1) und detaillierter Provenienzinformatio-nen für die Nachnutzbarkeit (R1.2).

Sobald für eine Ressource entschieden wird, die FAIR-Prinzipien einzuhalten,

sollte eine erste Evaluation erfolgen, um den aktuellen Grad der FAIRness festzustellen (sog. FAIR-Assessment) und dann im zweiten Schritt die Priorisierung für Verbesserungsschritte durchzuführen (FAIRification). Seit 2016 wurde eine Reihe von Werkzeugen und Richtlinien für beide Schritte entwickelt, sowohl für Daten als auch Infrastrukturkomponenten und basierend auf den unterschiedlichen, domänenspezifischen Interpretationen der FAIR-Prinzipien.

Feststellung der FAIRness einer Ressource: FAIR-Assessment

Ziel der 2013 gegründeten internationalen Research Data Alliance (RDA) ist es, den Datenaustausch über Technologien, Disziplinen und Länder hinweg zu ermöglichen [24]. Als eine der ersten Initiativen hat die RDA bereits 2019 eine Arbeitsgruppe zur Entwicklung eines Bewertungsschemas für die FAIRness von Daten gegründet. Dieses Schema wurde später als „FAIR Data Maturity Model“

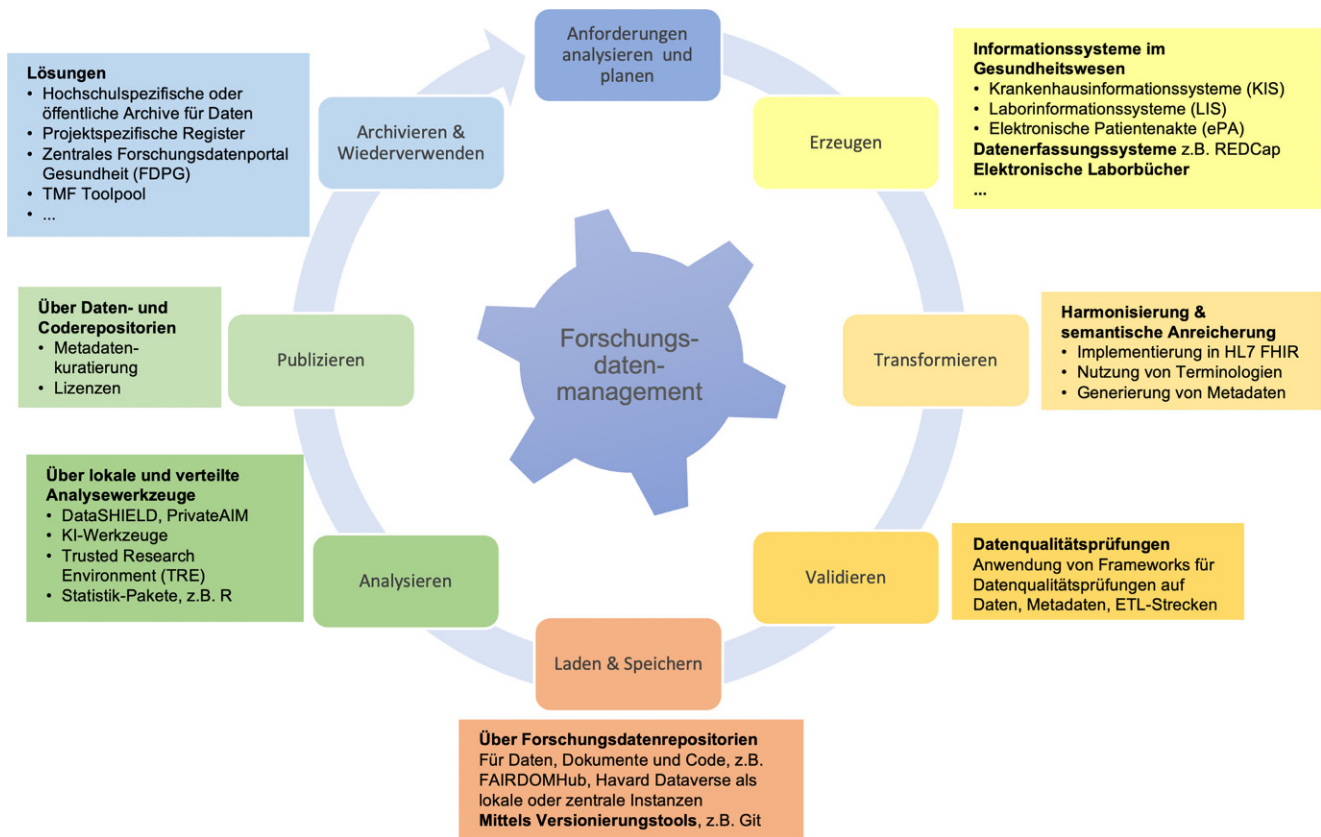


Abb. 1 ▲ Konzepte und Werkzeuge für FAIRes Forschungsdatenmanagement entlang des Datenlebenszyklus. In der Gesundheitsforschung werden heterogene Daten in dynamischen Infrastrukturen unter Einhaltung gesetzlicher Vorgaben verwaltet. Dies erfordert eine gezielte Planung und Umsetzung einheitlicher Strategien für das Datenmanagement entlang des Datenlebenszyklus. Planung befasst sich mit vorausschauenden Aspekten zu Durchführung und Dokumentation zum Sammeln, Transformieren, Qualitätsanalyse, statistische Analyse und Reporting von Gesundheitsdaten sowie deren Archivierung. Zentral ist hierbei die Auswahl geeigneter Metadaten und Metadatenstandards in einer handhabbaren Granularität und hoher Qualität. Beispielfhaft zeigt die Abbildung die Verarbeitung von Daten des MII-KDS mithilfe standardisierter ETL(extrahieren, transformieren, laden)-Entwicklungsprozesse entlang des Datenlebenszyklus². Der Datenlebenszyklus beschreibt den Weg der (Wieder-)Verwendung von Routinedaten von der Datensammlung bis zur endgültigen Analyse und Veröffentlichung unter Berücksichtigung (inter)nationaler Vorgaben, z. B. Datenschutz-Grundverordnung (DSGVO). Dabei erfolgt der Datentransfer mehrstufig mit dem Ziel, die Daten in interoperable Datenformate und -standards zu konvertieren. Methodisch gelingt dies unter Anwendung eines gezielten Datenmappings der Rohdatenwerte in einer temporären Speicherumgebung bzw. lokalen Data-Marts. *REDCap* Research Electronic Data Capture HL7; *FHIR* Healthcare Level 7 Fast Healthcare Interoperable Resources; *ETL* Extract, Transform, Load; *TMF* Technologie- und Methodenplattform für die vernetzte medizinische Forschung e. V.; *KI* Künstliche Intelligenz. (Quelle: eigene Abbildung). ²Gemäß <https://forschungsdaten.info/themen/informieren-und-planen/datenlebenszyklus/>, Zugriffen: 26. Februar 2024

[25] bekannt und gilt heute als ein zentraler Bewertungskatalog in vielen Wissenschaftsbereichen [26]. Ergebnis der Arbeit ist eine Liste von domänenunabhängigen FAIRness-Indikatoren mit einem zugehörigen Reifegradmodell. Basierend darauf entstanden verschiedene Richtlinien und Checklisten zur Implementierung der FAIR-Indikatoren [27]. Der Grad der FAIRness kann pro Indikator auf einer 5-stufigen Skala oder durch Zuweisung einer Ja/Nein-Bewertung angegeben werden („not applicable bis fully implemented“; [26]). Darüber hinaus un-

terstützen diverse Werkzeuge das FAIR-Assessment, beispielsweise manuelle Fragebögen, Checklisten oder automatisierte Tests.

Manuelle FAIR-Bewertungsmethoden in der Form von Self-Assessments bieten eine sehr einfache und schnelle Möglichkeit, die FAIRness eines Forschungsobjekts zu bewerten, werden jedoch in der Literatur als schwerfällig beschrieben [28]. Ein Beispiel für das FAIR-Assessment des DZD-Basisdatensatzes mit SATIFYD (Self-Assessment

Tool to Improve the FAIRness of Your Dataset) zeigt **Abb. 2**.

Halbautomatische Methoden ermöglichen es, die Bewertung teilweise automatisch vorzunehmen und die Ergebnisse anschließend zu korrigieren, zu verfeinern und zu vervollständigen [29]. Automatisierte Bewertungen werden in der Regel über Webanwendungen und entsprechende Schnittstellen (APIs) zugänglich gemacht. Das zu bewertende Forschungsobjekt wird in der Regel über einen Uniform Resource Identi-

Tab. 1 Anforderungen an Gute Forschungsdaten gemäß FAIR-Kriterienkatalog [11] am Beispiel ausgewählter Gesundheits(meta)daten: Die Anforderungen werden klassifiziert nach der Verbesserung der Auffindbarkeit (Findability), Erreichbarkeit (Accessibility), Interoperabilität (Interoperability) und Nachnutzbarkeit (Reusability). Spalten 2 und 3 der Tabelle zeigen exemplarisch positive Beispiele aus aktuellen Forschungsdatensätzen ausgewählter nationaler Infrastruktur- und Verbundprojekte. Die Tabelle erhebt keinen Anspruch auf Vollständigkeit

FAIR-Kriterium	Beispiel	Erklärung
<i>Findable</i>		
F1: Datenpunkte und Metadaten besitzen einen global eindeutigen und dauerhaft gültigen Identifier	Jedes Element des NFDI4Health-MDS ist durch einen Identifier identifiziert	Jedes Element des NFDI4Health-MDS ist mit einem eindeutigen Identifier innerhalb des MDS-Namensraums versehen, der einem bestimmten Schema folgt. Dadurch kann auf alle Metadaten eindeutig verwiesen werden
F2: Die Daten sind mit reichhaltigen Metadaten annotiert	–	–
F3: Metadaten müssen eindeutig und explizit den Identifier der beschriebenen Daten enthalten	Die eindeutige Beschreibung der Metadaten des zentralen Study-Hubs der NFDI4Health ist im MDS spezifiziert [14]	Zur Referenzierbarkeit von Metadaten (z. B. wenn sie getrennt von den eigentlichen Datensätzen abgelegt werden) ist es notwendig, dass Metadaten eindeutig identifizierbar sind
F4: (Meta-)Daten werden in einer durchsuchbaren Ressource registriert oder indiziert	Das FDPG bietet angemeldeten Nutzerinnen und Nutzern die Möglichkeit, verfügbare KDS-Daten über die angeschlossenen Datenintegrationszentren zu suchen	Eine Suche über Datenmodelle und deren Metadaten stellt sicher, dass Nutzerinnen und Nutzer darüber informiert werden, welche Daten über ein Repository bereitgestellt werden können und dass diese Daten auch gefunden werden
<i>Accessible</i>		
A1: (Meta-)Daten sind anhand ihres Identifier über ein standardisiertes Kommunikationsprotokoll abrufbar	Die Datenpunkte des MII-KDS sind über eindeutige Identifier über https referenzierbar	Der diagnostische Laborbefund ist erreichbar über https://www.medizininformatik-initiative.de/fhir/core/modul-labor/StructureDefinition/DiagnosticReportLab
A1.1: Das Protokoll ist offen und universell implementierbar	Datenherausgabeprozess über das FDPG	Für die Datenherausgabe über das FDPG existiert ein protokollierter Standardprozess
A1.2: Das Protokoll unterstützt ein Authentifizierungs- und Autorisierungsverfahren, falls erforderlich	–	–
A2: Metadaten sind auch zugänglich, wenn die dazugehörigen Daten nicht verfügbar sind	Das Datenmodell des DZD-KDS ist über das MDM-Portal [12] öffentlich zugänglich, ohne dass die zugehörigen Daten verfügbar sind	Potenzielle Datennutzerinnen und -nutzer sowie Kooperationspartnerinnen und -partner können sich über das MDM bereits über die am DZD verfügbaren Datenpunkte informieren, ohne selbst Zugriff auf die Daten zu erhalten. Es besteht ebenso die Möglichkeit, Datenmodelle über Wissensgraphen miteinander zu verknüpfen, z. B. [20]
<i>Interoperable</i>		
I1: (Meta-)Daten verwenden eine formale, zugängliche, gemeinsame und breit anwendbare Sprache zur Wissensdarstellung	Der MII-KDS ist im FHIR-Format implementiert (z. B. Modul Person, https://simplifier.net/MedizininformatikInitiative-ModulPerson/)	Die Verwendung von FHIR garantiert, dass Menschen (und Maschinen) die Daten lesen, (gleichartig) interpretieren und miteinander über verschiedene Anwendungssysteme hinweg austauschen können, z. B. über ein verteiltes Anfragesystem [21]
I2: (Meta-)Daten nutzen Vokabulare, die den FAIR-Prinzipien folgen	Der MII-KDS nutzt LOINC zur Annotation der Laborparameter	LOINC selbst kann als FAIRes Vokabular betrachtet werden. Somit wird sichergestellt, dass die Vokabulare, die genutzt werden, um die Daten zu beschreiben, selber zugänglich und verständlich sind
I3: (Meta-)Daten enthalten qualifizierte Referenzen zu anderen (Meta-)Daten	Metadaten im HL7-FHIR-Standard verlinken auf weitere (FHIR-)Ressourcen und kontrollierte Vokabulare	Beispielsweise verweisen Abrechnungsdaten aus dem MII-KDS auf ICD-10-GM-Kodierungen (erforderlich laut §§ 301, 295, 5. Sozialgesetzbuch). Eine Vielzahl solcher qualifizierten Referenzen auf andere Vokabulare ist notwendig, um ein umfangreiches Bild der Daten zu erhalten [3]
<i>Reusable</i>		
R1: (Meta-)Daten sind detailliert beschrieben und enthalten präzise, relevante Attribute	–	–
R1.1: (Meta-)Daten enthalten eine eindeutige zugreifbare Angabe einer Nutzungslizenz	License DZD-KDS: Creative Commons BY-SA 4.0	Der DZD-KDS ist mit einer offenen Standardlizenz der Creative Commons versehen. Die Lizenz ist Teil der Metadaten. Sie klärt eindeutig die Bedingungen zur Nachnutzung eines Datensatzes [16]
R1.2: (Meta-)Daten enthalten detaillierte Provenienzinformationen	–	Wünschenswert wäre die Implementierung eines W3C-Standards für Provenienzinformation über die KDS-Elemente [22]
R1.3: (Meta-)Daten entsprechen fachgebietsrelevanten Standards	Repräsentation des MII-KDS als FHIR-Ressourcen	Die Verwendung des internationalen HL7-FHIR-Standards erhöht die Interoperabilität der MII-Ressourcen mit Software, aber beispielsweise auch mit dem MDS der NFDI4Health, welches ebenfalls mit FHIR-Ressourcen ^a abgebildet wurde [23]

^aNFDI4Health Nationale Forschungsdateninfrastruktur für personenbezogene Gesundheitsdaten, MDS Metadatenschema, FDPG Forschungsdatenportal für Gesundheit, KDS Kerndatensatz, DZD Deutsches Zentrum für Diabetesforschung, MDM Medical Data Model, MII Medizininformatik-Initiative, FHIR Fast Healthcare Interoperability Resources, LOINC Logical Observation Identifiers Names and Codes, HL7 Health Level 7, ICD-10-GM International Classification of Diagnoses 10. Revision German Modification, W3C World Wide Web Consortium
^b<https://simplifier.net/NFDI4Health-Metadata-Schema/>, Zugegriffen: 26. Februar 2024

fier (URI) oder einen Digital Object Identifier (DOI) bereitgestellt.

Für das FAIR-Assessment im medizinischen und Gesundheitsbereich spielen die semiautomatischen Verfahren eine entscheidende Rolle, da insbesondere die Einschätzung des Informationsgehalts von Metadaten eine manuelle Bewertung erfordert. Andere Aspekte können sehr gut automatisiert bewertet werden, beispielsweise die Feststellung

der Maschinenlesbarkeit [30]. Der Einsatz halbautomatischer Bewertungstools erfordert jedoch Übung im Umgang und Kenntnisse darüber, wie sich die jeweiligen Indikatoren auf die Daten und FDM-Praktiken auswirken [26]. Vorschläge für entsprechende Werkzeuge sind in **Tab. 2** zusammengefasst [31–38].

Workflows zur Verbesserung der FAIRness: FAIRification

Sobald ein FAIR-Assessment für einen Datensatz vorliegt, kann ein strukturierter Prozess zur Verbesserung der FAIRness gestartet werden. Ziel dieser FAIRification ist es, in den 4 Teilbereichen (F, A, I, R) eine verbesserte Wertung (Score) zu erhalten. Hierbei wird in der Regel nicht auf die 100-prozentige Er-

Tab. 2 Auswahl an FAIR-Assessment-Werkzeugen. Es wird unterschieden nach manuellen (m), semiautomatischen (s) und automatischen (a) Werkzeugen für das FAIRness-Assessment von Ressourcen, insbesondere (Meta-)Datenspezifikationen

Werkzeug	Kurzbeschreibung	Typ
FAIR Data Maturity Model (FDMM; [26])	Definition einer Menge von Indikatoren für FAIRness mit verschiedenen Implementierungsstufen	m
SHARC [31]	Assessment-Tool basierend auf den Regeln und Prinzipien der Web Content Accessibility Guidelines ^a	m
Ontology FAIRness Evaluator (O'FAIRe; [32])	Werkzeug zur automatischen Bewertung der FAIRness einer semantischen Ressource oder Ontologie	a
FOOPS [33]	Online-Dienst zur Bewertung der Konformität von Vokabularen oder Ontologien mit den FAIR-Grundsätzen	a
F-UJI [34]	Automatisches Online-Assessment-Tool basierend auf von FAIRsFAIR entwickelten Metriken	a
10 simple rules for making a vocabulary FAIR [35]	Anleitung mit 10 Schritten um ein Vokabular FAIR zu machen	m
ARDC FAIR self-assessment [36]	Kurzer Fragebogen mit hilfreichen Erklärungen für jede Frage	m
FAIRshake [37]	Assessment-Tool für digitale Objekte	s
FAIR-Aware [38]	Fragebogen mit 10 Fragen und hilfreichen Erklärungen	m

^a<https://www.w3.org/TR/WCAG20/>, Zugriffen: 26. Februar 2024

füllung aller Kriterien abgezielt, sondern es werden die projektspezifischen Anforderungen analysiert und bestimmte Kriterien entsprechend priorisiert. Beispielsweise sollte ein Metadatenschema vor allem zugreifbar und interoperabel sein, ein Datensatz eher auffindbar und interoperabel, ein Provenienzmodell vor allem nachnutzbar. Interoperabilität hat im Allgemeinen einen besonderen Stellenwert für die Wiederverwendung von Daten im Gesundheitswesen. Im medizinischen Kontext beschreibt HL7 FHIR Datenformate und Elemente als sogenannte Ressourcen und bietet Schnittstellen an, um diese auszutauschen. Die Vorteile der etablierten HL7-Standardproduktlinien Version 2, Version 3 und Clinical Document Architecture (CDA) werden dabei mit jenen aktueller Webstandards kombiniert. Ein starker Fokus liegt dabei auf einer einfachen Implementierbarkeit, um den Datenaustausch zwischen Softwaresystemen im Gesundheitswesen zu befördern.

Auch FAIRification wird durch verschiedene Werkzeuge unterstützt, die entweder generisch oder domänenspezifisch gehalten sind [39]. Speziell an die Anforderungen von Gesundheitsforschungsdaten angepasste Workflows berücksichtigen vor allem die technischen, ethischen und rechtlichen Anforderun-

gen [40]. Das FAIRplus-Konsortium⁵ hat hierfür ein umfassendes Rahmenwerk entwickelt [41], welches frei von spezifischen Implementierungslösungen oder -methoden verwendbar ist. Es besteht aus 4 verschiedenen Phasen: einer Zieldefinitionsphase, einer anfänglichen Projektprüfungsphase, einer iterativen zyklischen FAIRifizierungsphase und einer Post-FAIRifizierungsüberprüfung. Ein weiterer FAIRification-Workflow für Gesundheitsdaten wurde vom FAIR4Health-Projekt entwickelt und evaluiert [40, 42].

Das DZD strebt eine FAIRe Datenverwaltung auf Basis des DZD-KDS an, um die Nutzung der Daten innerhalb des DZD und der breiteren Diabetes-Forschungsgemeinschaft zu unterstützen. Im Rahmen eines strukturierten, begleitenden FAIRification-Prozesses wurde beispielsweise der DZD-KDS in der Entwicklungsphase bereits mit Metadaten angereichert und im Medical-data-model(MDM)-Portal registriert. Er steht in mehreren maschinen- und menschenlesbaren Formaten zur Verfügung. Mit einer offenen Lizenz trägt er zur Standardisierung in der Erfassung von Stoffwechselfparametern in der klinischen Forschung bei. Das Ergebnis des

⁵ <https://fairplus-project.eu>, Zugriffen: 26. Februar 2024.

manuellen FAIR-Assessments nach der FAIRification ist in **Abb. 2** abgebildet. Weitere Beispiele für bereits erfolgte FAIR-Assessments mit partieller FAIRification sind das NFDI4Health-Metadatenschema, die LOINC-Kodierung innerhalb der SHIP(Study of Health in Pomerania)-Studie, die Datensätze der CODEX+(COVID-19 Data Exchange Platform)-Projekte oder biomedizinische Simulationsmodelle.

FAIR betrifft alle Bausteine des Forschungsdatenmanagements

Die FAIR-Prinzipien werden häufig im Kontext von (Forschungs-)Daten betrachtet, jedoch wird bereits in der Originalpublikation darauf hingewiesen, dass sie grundsätzlich auf alle digitalen Artefakte in einem Forschungsprozess angewendet werden sollen, also auch auf Software, Analyseskripte oder auch Ausführungsumgebungen. Im Kontext der MII betrifft dies z. B. die Implementierung der ETL-Prozesse in den Datenintegrationszentren, aber man kann dies grundsätzlich auch auf die Informations- und Entscheidungsunterstützungssysteme der Krankenversorgung beziehen, die sowohl als Quell- als auch als Zielsysteme insbesondere in der versorgungsnahen Forschung dienen können. Im Prinzip lassen sich viele der oben genannten FAIR-Kriterien direkt auf Software und Ausführungsumgebungen anwenden, wie z. B. die Bereitstellung von Software über Plattformen wie codeberg, mit eindeutigen Identifier für verwendete Softwareversionen oder die Bereitstellung von Ausführungsumgebungen als virtuelle Maschinen oder Container [43]. Allerdings existieren bisher kaum Metadatenstandards für die Beschreibung von Software und Ausführungsumgebungen [44, 45]. Auch für die Nutzbarkeit von Versorgungsdaten, beispielsweise in KI-Anwendungen, können die FAIR-Kriterien herangezogen werden, um die Reproduzierbarkeit von Modellen zu erhöhen. Die Abdeckung der FAIR-Kriterien in Guidelines für KI-Modellentwicklung variiert hierbei stark – weitere Forschung ist auf diesem Gebiet erforderlich [46].

The image displays four panels of a FAIR assessment tool, each with a large blue letter at the top and a progress indicator. The panels are:

- FINDABLE:** 100% Well done. Questions include: "Did you provide sufficient metadata (information) about your data for others to find, understand and reuse your data?" (Rich metadata with as much information as possible), "Did you use standards such as controlled vocabularies, taxonomies (thesauri) or ontologies to describe your dataset?" (Controlled vocabularies, Taxonomies (thesauri), Ontologies, There are no standards for my discipline), and "Did you provide rich and detailed additional documentation?" (Readme file, Versioning, Provenance).
- ACCESSIBLE:** 100% Well done. Questions include: "Is the metadata publicly accessible even if the data is no longer available?" (Yes, No, I can't find this information in EASY), "Does your dataset contain personal data?" (Yes, No), and "Which of the usage licenses provided by EASY did you choose in order to comply with the access rights attached to the data?" (Open access (CC0)).
- INTEROPERABLE:** 92% Want to improve?. Questions include: "Are the data in your dataset stored in preferred formats?" (All of the data are in preferred formats), "Do you link to other (meta)data and is this (meta)data online resolvable?" (Yes, and these are online accessible), and "Did you provide contextual information about your dataset?" (Persistent Identifier(s), Reference to other datasets, Reference to publications, No contextual metadata).
- REUSABLE:** 100% Well done. Questions include: "What kind of information did you provide about the provenance of your data?" (Origin of data, Citations for reused data, Workflow description for collecting data (machine readable), Processing history of data, Version history of data), "Which usage license provided by EASY did you choose for your dataset?" (Open access (CC0)), and "Does your (meta)data meet domain standards?" (Domain standards in metadata).

Abb. 2 ▲ Ergebnisse des FAIR-Assessments des Basisdatensatzes des Deutschen Zentrums für Diabetesforschung (DZD) nach Umsetzung der Empfehlungen des SATIFYD-Tools. Im Rahmen eines strukturierten FAIRification-Assessments wurden folgende Schritte auf die ursprüngliche Version des Kerndatensatzes (KDS) angewendet: Zuweisung einer dauerhaften Kennung sowohl für die Daten als auch für die Metadaten, Versionierung, Konvertierung in das empfohlene Format für Tabellenkalkulationen, Annotation der darin enthaltenen Parameter mit dem Unified Medical Language System (UMLS), Lizenzierung des Datensatzes, Indizierung der Metadaten als eine durchsuchbare Ressource, die den Datensatz großzügig mit Metadaten anreichert und Herkunftsinformationen bereitstellt. Diese Maßnahmen führten zu einem FAIRification-Score von 98 % im Online-Assessment-Tool SATIFYD³ (Quelle: eigene Abbildung). ³<https://satifyd.dans.knaw.nl/>, Zugegriffen: 26. Februar 2024

Die gemeinsame (Nach-)Nutzung von Ausführungsumgebungen als sogenannte vertrauenswürdige Forschungsumgebungen (Trusted Research Environments; [47]), die die Herausgabe von Gesundheitsdaten vermeiden, geht darüber hinaus und erfordert insbesondere Maßnahmen der sicheren Nutzung durch verschiedene Anwenderinnen und Anwender. Auch solche Umgebungen können im Rahmen von FAIR-Kriterien beschrieben werden, so z. B. Interoperabilität der Prozesse und Nutzungsdaten, definierte Nutzungsbedingungen und Verwendung von offenen Kommunikationsstandards [46]. Andere Anforderungen wie Skalierbarkeit und Sicherstellung von Vertrauenswürdigkeit gehen darüber hinaus.

Kosten vs. Nutzen von FAIRification

Eine der größten Herausforderungen bei dem Bestreben, Daten und Infrastrukturen FAIR zu gestalten – und vielleicht der digitalen Transformation insgesamt – besteht darin, dass diejenigen, die von der Nachnutzung der Daten profitieren,

oftmals andere Personen sind als diejenigen, die sie erheben und aufbereiten. Der Konflikt zwischen diesen beiden Interessengruppen wird bei Gesundheitsdaten sehr deutlich. Quellsysteme wie Krankenhausinformationssysteme und elektronische Patientenakten basieren meist auf Freitext, der nur sehr begrenzt strukturiert ist und wichtige Datenelemente nur teilweise erfasst. Dem medizinischen Fachpersonal steht nur eine begrenzte Zeit für die Dateneingabe zur Verfügung. Andererseits können geringe Qualität, unvollständige und unstrukturierte Daten im Quellsystem nicht durch FAIRification korrigiert werden. Um Daten effektiv und in der Breite wiederverwenden zu können, wie z. B. durch Forschende, aber auch durch medizinisches Fachpersonal, Patientinnen, Patienten und Industrie, muss in die Quellsysteme investiert werden, um die Datenqualität und FAIRness am Ort der Datenerhebung zu verbessern.

Unter diesem Gesichtspunkt ist es natürlich wichtig, die Perspektive der Datenproduzentinnen und -produzenten zu berücksichtigen. Im Gesundheitswesen ist dies üblicherweise das medizinisch-

pflegerische Fachpersonal. Es muss diskutiert werden, in welchem Verhältnis die Kosten für die FAIRere Datenbereitstellung zu den jeweiligen Vorteilen stehen. Kosten können beispielsweise durch eine bessere Digitalisierung gesenkt werden, aber auch durch Prozessoptimierungen und Reduktion doppelter und grundsätzliche Vereinfachung von Dateneingaben. Weiterhin können die Vorteile guter Daten direkt an die Datenerzeuger zurückgespielt werden, beispielsweise indem offene Werkzeuge bereitgestellt werden, um Patientendaten zu analysieren, mit ähnlichen Patientinnen und Patienten (auch aus anderen Einrichtungen) zu vergleichen und mit visueller Unterstützung die Daten der jeweiligen Gesundheitseinrichtung zu analysieren oder zu explorieren. Auf diese Weise werden die Datenproduzentinnen und -produzenten zeitnah zu denjenigen, die von den aufbereiteten, semantisch angereicherten FAIReren Daten profitieren.

Anknüpfung an nationale und internationale Initiativen

Viele europäische Nachbarländer haben in den letzten Jahren in die Entwicklung eines nationalen Rahmens für semantische Interoperabilität von Gesundheitsdaten investiert. So wurden Konzepte für eine umfassende Datenanreicherung entwickelt, um die Datenanalyse zu vereinfachen sowie die korrekte Interpretierbarkeit der Daten bei deren Weiterverwendung in verschiedenen Kontexten zu erhalten. Dies zumeist mit dem Ziel, nachhaltige und skalierbare Prozesse und Strukturen zu entwickeln, die eine zeit- und kosteneffiziente Bereitstellung von standardisierten, FAIRen Gesundheitsdaten für die Primär- und Sekundärnutzung gewährleisten.

Die nationalen Initiativen, die vielerorts länderübergreifend in engem Austausch stehen, entwickeln zumeist individuelle, unabhängige Informationsmodelle, welche die Bereitstellung harmonisierter Daten für Forschungszwecke ermöglichen. Dabei beziehen sich alle auf international etablierte Kodiersysteme, Ontologien und Terminologien, wie beispielsweise ICD, SNOMED oder LOINC. Die nationalen Forschungsdateninfrastrukturen (NFDIs) in Deutschland tätigen wichtige fachübergreifende Investitionen in die Implementierung der FAIR-Kriterien. Im Gesundheitsbereich sind hierbei das Deutsche Humangenom-Phänomarchiv⁶ (GHGA), die NFDI4Immuno⁷ und NFDI4Health⁸ als Beispiele zu nennen.

Selbstverständlich existieren auch länderspezifische Standards, die aufgrund der Kodierungsprozesse für die Vergütung von Gesundheitsleistungen in den Krankenhäusern verwendet und im Rahmen der Initiativen entsprechend genutzt werden. Während der modulare MII-KDS die Struktur, das Format und die semantische Annotation der MII-Daten auf der Grundlage von HL7 FHIR

beschreibt, setzt beispielsweise die Swiss Personalized-Health-Network-(SPHN-) Initiative⁹ mit ihrer Interoperabilitätsstrategie auf die semantische Darstellung von Gesundheitsdaten in einem „Knowledge Graphen“, unter Anwendung der „Semantic Web Technologie“ und stringenter Umsetzung der FAIR-Prinzipien, um die gemeinsame Nutzung und nahtlose Integration unterschiedlicher gesundheitsbezogener Daten zu erleichtern [48]. Ein konkretes Beispiel für die länderübergreifende FAIRification von Gesundheitsdaten ist die gemeinsame Arbeit an einem Provenienzmodell für die KDS der MII und SPHN im Rahmen eines Community-getriebenen Projekts¹⁰ (MInimal Requirements for Automated Provenance Information Enrichment [MIRAPIE]).

Die Interoperabilitätsstrategie der niederländischen Initiative „Health-RI“ befindet sich derzeit noch in der Planung und soll in einem sogenannten Blütenmodell organisiert werden, bei welchem der KDS unter Nutzung international anerkannter Datenstandards zentral definiert wird und die verschiedenen Blütenblätter individuelle Projekt- oder Bereichsschemata darstellen. Das FAIR-Datenimplementierungsteam von Health-RI entwickelt zusammen mit den Netzwerkpartnern einen Ansatz für die FAIRE Datenimplementierung, d.h. die Bereitstellung von Datensätzen und zugehörigen (reichhaltigen) Metadaten in allen angeschlossenen Organisationen, welche Daten verwalten. Ziel ist es, Prozesse zur FAIRification so früh wie möglich im Datenbereitstellungsprozess anzugehen und weitestgehend zu automatisieren. In mehreren skandinavischen Ländern, aber beispielsweise auch in Katalonien kommt openEHR¹¹ zum Einsatz. openEHR bietet die Möglichkeit, eine FAIR-konforme klinische Datenressource aufzubauen, indem die openEHR-Spezifikationen in Kombination mit eini-

gen Ad-hoc-Einsatzkonfigurationen angewendet werden [49].

Die nationalen und länderübergreifenden Dateninitiativen der europäischen Länder spielen eine entscheidende Rolle bei der Förderung der Weiterverwendung klinischer Gesundheitsdaten sowie biologischer Humanen im Rahmen des Europäischen Gesundheitsdatenraums¹² (EHDS). Die Vorarbeiten dieser Initiativen mit Blick auf FAIRE Gesundheitsdaten erfordern, dass auch auf europäischer Ebene eine harmonisierte und interoperable Dateninfrastruktur geschaffen wird, die den Austausch und die Weiterverwendung von Gesundheitsdaten für die Forschung erleichtert. Auf internationaler Ebene spielt die Interaktion mit Observational Health Data Sciences and Informatics (OHDSI) eine zunehmend wichtige Rolle. Die OHDSI-Community entwickelt Methoden zur Analyse von Beobachtungsdaten und wendet diese in konkreten Projekten auf große Datenmengen an. Darüber hinaus fördert OHDSI die offene Zusammenarbeit und den Austausch von Ideen, Methoden und Ergebnissen, beispielsweise durch Code-Repositories, Tutorien und Workshops. Dies erleichtert die Interoperabilität und die Anwendung der FAIR-Prinzipien, indem es eine gemeinsame Struktur für verschiedene Arten von Gesundheitsdaten bereitstellt. In der MII und dem NUM (Projekt CODEX+) wurden bereits OHDSI-basierte Pakete entwickelt und bereitgestellt, die es den Uniklinika ermöglichen, an internationalen Studien teilzunehmen.

Fazit

Forschung an Gesundheitsdaten führt zu neuen medizinischen Erkenntnissen, unterstützt Therapieentscheidungen, trägt zur Entwicklung neuartiger technischer und KI-basierter Entwicklungen für Datenerhebung und Datenanalysen bei und verbessert somit die Patientenversorgung. Hierbei sind viele Forschungsfragen auf nationaler und internationaler Ebene im Verbund zu adressieren. Das

⁶ <https://www.ghga.de/>, Zugriffen: 26. Februar 2024.

⁷ <https://www.nfdi4immuno.de/>, Zugriffen: 26. Februar 2024.

⁸ <https://www.nfdi4health.de/>, Zugriffen: 26. Februar 2024.

⁹ <https://www.sphn.ch>, Zugriffen: 26. Februar 2024.

¹⁰ <https://codeberg.org/MIRAPIE/>, Zugriffen: 26. Februar 2024.

¹¹ <https://www.openehr.org/>, Zugriffen: 26. Februar 2024.

¹² <https://www.european-health-data-space.com>, Zugriffen: 26. Februar 2024.

Teilen qualitätsgesicherter, vollständiger und gut dokumentierter Daten aus Versorgung und Forschung ist hierbei essenziell. Dies umfasst auch zunehmend genetische Daten und von Patientinnen und Patienten bereitgestellte Informationen.

Voraussetzung für alle Datennutzungsprozesse sind gute Planung und kontinuierliches Monitoring. Ein modernes FDM erhöht die Datenqualität, macht Ergebnisse transparent und unterstützt somit gute Forschung. Die FAIR-Prinzipien gelten dabei als ein De-facto-Standard für das FDM, denn sie schaffen insbesondere in Kombination mit Provenienzmodellen, Guter Wissenschaftlicher Praxis und der Umsetzung von Prinzipien für gute Software und Forschungsinfrastrukturen Vertrauen. Sie motivieren somit zur Nachnutzung vorhandener Daten.

Ein Grundmaß an FAIRness sollte für jedes Datenprojekt im Gesundheitssektor das Ziel sein, eine vollständige FAIRification ist wahrscheinlich nur in wenigen Projekten von praktischer Relevanz. Es ist daher essenziell, dass sich Forschende mit der Planung eines zielgerichteten FDM auseinandersetzen, insbesondere aber auch die Datenerhebenden in den Prozess einbezogen werden und Kompetenzen und Ressourcen entsprechend deren zentraler Rolle an den Datenquellen berücksichtigt werden. Neben der Nutzung an den Forschungseinrichtungen selbst, beispielsweise die Inanspruchnahme von Beratungsangeboten durch Data Stewards oder Weiterbildungen an Graduiertenakademien und Datenintegrationszentren, stehen auch auf nationaler und internationaler Ebene zahlreiche Angebote und Handbücher zur Verfügung. Beispielhaft seien die Landesinitiativen FDM¹³, die NFDI4Health oder Initiativen auf europäischer Ebene (ELIXIR¹⁴, Research Data Alliance) genannt.

Forschende sollten im Rahmen der gesetzlichen Möglichkeiten die zur Pu-

blikation gehörigen Datensätze und Metadaten publizieren. Wie die Daten nachgenutzt werden und in welcher Form die Datenproduzenten hierbei mit einbezogen werden sollen, klärt die Lizenzierung, deren Angabe Teil der FAIR-Anforderungen ist. Vertrauenswürdige, domänenspezifische Repositorien, wie beispielsweise der zentrale Study-Hub der NFDI4Health oder das FDGP für Routinedaten mit Broad Consent, stellen die publizierten Daten und Metadaten anschließend über koordinierte Prozesse für die Nachnutzung zur Verfügung. Die derartige Nachnutzung von Forschungsdaten garantiert in der Regel eine hohe Datenqualität und reduziert die Aufwände für Personal, IT-Infrastruktur und Zeit gegenüber der Erhebung von Primärdaten. Zu bedenken gibt es hierbei jedoch, dass die Datenqualität in den primären Erhebungssystemen bereits maßgeblich über die Qualität der daraus abgeleiteten Forschungsdaten bestimmt – somit also auch ein Umdenken in der Routinedatenerhebung erfolgen muss.

Gesundheitsdaten werden in ganz Europa auf unterschiedliche Weise erhoben. Nur wenn Gesundheitsdaten FAIR sind, können sie neben der lokalen und nationalen Nutzung auch an internationale Netzwerke anknüpfen. So wurden beispielsweise im Rahmen des EHDS Standards festgelegt, die sicherstellen sollen, dass Daten grenzüberschreitend für Forschung und Entscheidungsfindung genutzt werden können.¹⁵

Korrespondenzadresse

Prof. Dr. Dagmar Waltemath

Abteilung Medizininformatik, Institut für Community Medicine
Walther-Rathenau-Straße 48, 17475 Greifswald, Deutschland
dagmar.waltemath@med.uni-greifswald.de

Förderung. Die Arbeit von E.T. Inau wird aus dem DFG-Projekt NFDI4Health finanziert. Die Arbeit von P. Gröber wird aus dem BMBF NUM-DIZ (01KX2121) finanziert. L. Michaelis wird aus dem EOSC-FAIR-IMPACT-Projekt und der BMBF-Nachwuchsgruppe MeDaX (01ZZ2019) finanziert. M. Sedlmayr wird durch die BMBF-Programme MIRACUM und CODEX+

¹⁵ <https://tehdas.eu/app/uploads/2022/02/tehdas-identification-of-relevant-standards-and-data-models-for-semantic-harmonization-2022-02-03.pdf>, Zugriffen: 26. Februar 2024.

sowie durch das EHDEN Data Partner Programm finanziert. K. Crameri wird durch das Swiss State Secretariat for Education, Research and Innovation finanziert. D. Krefting ist an den Projekten CAEHR (01ZZ2103A), basetrace (01ZZ2306A), MII HiHmed (01ZZ1802B) und ACRIBIS (01ZZ2317B) beteiligt.

Funding. Open Access funding enabled and organized by Projekt DEAL.

Einhaltung ethischer Richtlinien

Interessenkonflikt. D. Waltemath, O. Beyan, K. Crameri, A. Dedié, K. Gierend, P. Gröber, E. T. Inau, L. Michaelis, I. Reinecke, M. Sedlmayr, S. Thun und D. Krefting geben an, dass kein Interessenkonflikt besteht.

Für diesen Beitrag wurden von den Autor/-innen keine Studien an Menschen oder Tieren durchgeführt. Für die aufgeführten Studien gelten die jeweils dort angegebenen ethischen Richtlinien.

Open Access. Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

Literatur

1. Semler SC, Wissing F, Heyder R (2018) German medical informatics initiative. *Methods of information in medicine*. 57(S 01):e50–6. <https://doi.org/10.3414/ME18-03-0003>
2. Miederer I, Rogasch JM, Fischer R et al (2023) The Medical Informatics Initiative and the Network University. *Medicine, Perspectives for Nuclear Medicine*. *Nuklearmedizin-NuclearMedicine* <https://doi.org/10.1055/a-2067-7642>
3. Muzoor MR, Schaarschmidt M, Krefting D, Oehm J, Riepenhausen S, Thun S (2021) Towards FAIR Patient Reported Outcome: Application of the Interoperability Principle for Mobile Pandemic Apps. *Stud Health Technol Inform* 287:85–86. <https://doi.org/10.3233/shti210820>
4. Brammen D, Greiner F, Kulla M et al (2022) AKTIN—The German Emergency Department Data Registry—real-time data from emergency medicine: Implementation and first results from 15 emergency departments with focus on Federal Joint Committee's guidelines on acuity

¹³ <https://forschungsdaten.info/fdm-im-deutschsprachigen-raum>, Zugriffen: 26. Februar 2024.

¹⁴ <https://elixir-europe.org/>, Zugriffen: 26. Februar 2024.

- assessment. *Medizinische Klinik. Intensivmed.* <https://doi.org/10.1007/s00063-020-00764-2>
5. Heyder R, Coordination Office NUM, NUKLEUS Study Group et al (2023) Das Netzwerk Universitätsmedizin: Technisch-organisatorische Ansätze für Forschungsdatenplattformen. *Bundesgesundheitsblatt* 66:114–125. <https://doi.org/10.1007/s00103-022-03649-1>
 6. Koch M, Richter J, Hauswaldt J, Krefting D (2023) How to Make Outpatient Healthcare Data in Germany Available for Research in the Dynamic Course of Digital Transformation. *Stud Health Technol Inform* 307:12–21. <https://doi.org/10.3233/shti230688>
 7. Hoffmann J, Hanß S, Kraus M et al (2023) The DZHK research platform: maximisation of scientific value by enabling access to health data and biological samples collected in cardiovascular clinical studies. *Clin Res Cardiol* 112:923–941. <https://doi.org/10.1007/s00392-023-02177-5>
 8. Schons M, Pilgram L, Reese JP et al (2022) The German National Pandemic Cohort Network (NAPKON): rationale, study design and baseline characteristics. *Eur J Epidemiol* 37(8):849–870. <https://doi.org/10.1007/s10654-022-00896-z>
 9. Hoffmann K, Pelz A, Karg E et al (2023) Data integration between clinical research and patient care: A framework for context-depending data sharing and in silico predictions. *Plos Digit Health* 2(5):e140. <https://doi.org/10.1371/journal.pdig.0000140>
 10. Gierend K, Freiesleben S, Kadioglu D, Siegel F, Ganslandt T, Waltemath D (2023) The Status of Data Management Practices Across German Medical Data Integration Centers: Mixed Methods Study. *J Med Internet Res* 25:e48809. <https://doi.org/10.2196/48809>
 11. Wilkinson MD, Dumontier M, Aalbersberg IJ et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*. 15:3(1):1–9. <https://doi.org/10.1038/sdata.2016.18>
 12. Dugas M, Neuhaus P, Meidt A et al (2016) Portal of medical data models: information infrastructure for medical research and healthcare. *Database* 2016:bav121. <https://doi.org/10.1093/database/bav121>
 13. Fluck J, Lindstädt B, Ahrens W et al (2021) NFDI4Health-Nationale Forschungsdateninfrastruktur für personenbezogene. Gesundheitsdaten Bausteine Forschungsdatenmanagement 2021(2):72–85. <https://doi.org/10.17192/bfdm.2021.2.8331>
 14. Abaza H, Shutsko A, Golebiewski M et al (2023) Metadata schema of the NFDI4Health and the NFDI4Health Task Force COVID-19 (V3_2). *Publisso*. <https://doi.org/10.4126/FRL01-006453422>
 15. Darms J, Henke J, Hu X, Schmidt CO, Golebiewski M, Fluck J (2021) Improving the FAIRness of Health Studies in Germany: The German Central Health Study Hub COVID-19. *Stud Health Technol Inform*. <https://doi.org/10.3233/shti210818>
 16. German Center for Diabetes Research (DZD). (2023) DZD Core Data Set—Metadata and SOP (1.1.1.). <https://doi.org/10.5281/zenodo.8090451>. Zugegriffen: data set
 17. Stäubert S, Ammon D, Winter A (2022) SMITH-DIZ-Referenzarchitektur – Methodik und Ergebnisse. SMITH. Day, Bd. 2022. Science, Aachen <https://doi.org/10.3205/22smith29>
 18. Tahar K, Verbuechel R, Martin T, Graessner H, Krefting D (2023) Local Data Quality Assessments on EHR-Based Real-World Data for Rare Diseases. *Stud Health Technol Inform* 302:292–296. <https://doi.org/10.3233/shti230121>
 19. Draeger C, Tute E, Schmidt CO et al (2023) Identifying Relevant FHIR Elements for Data Quality Assessment in the German Core Data Set. *Stud Health Technol Inform* 302:272–276. <https://doi.org/10.3233/shti230117>
 20. Menzel F, Waltemath D, Henkel R (2023) Exploring New Possibilities for Research Data Exploration Using the Example of the German Core Data. Set Caring Is Sharing—exploiting Value Data Health Innov. <https://doi.org/10.3233/shti230255>
 21. Gruendner J, Deppenwiese N, Folz M et al (2022) The Architecture of a Feasibility Query Portal for Distributed COVID-19 Fast Healthcare Interoperability Resources (FHIR) Patient Data Repositories: Design and Implementation Study. *JMIR Med Inform* 10(5):e36709. <https://doi.org/10.2196/36709>
 22. Gierend K, Krüger F, Genehr S et al (2023) Capturing provenance information for biomedical data and workflows: A scoping review. Preprint. *ResearchSquare*. <https://doi.org/10.21203/rs.3.rs-2408394/v1>
 23. Vorisek CN, Lehne M, Klopfenstein SA et al (2022) Fast healthcare interoperability resources (FHIR) for interoperability in health research: systematic review. *JMIR Med Inform* 10(7):e35724. <https://doi.org/10.2196/35724>
 24. Emergence, R. D. A. (2014). Guest Editorial Building Global Infrastructure for Data Sharing and Exchange Through the Research Data Alliance. *D-Lib Magazine*, 20(1/2). <https://doi.org/10.1045/january2014-berman>
 25. Bahim C, Casorrán-Amilburu C, Dekkers M et al (2020) The FAIR Data Maturity Model: An Approach to Harmonize FAIR Assessments. *Data Sci J*. <https://doi.org/10.5334/dsj-2020-041>
 26. FAIR Data Maturity Model Working Group. (2020) FAIR Data Maturity Model. *Specif Guidel*. <https://doi.org/10.15497/rda00050>
 27. Bahim C, Dekkers M, Wyns B (2019) Results of an Analysis of Existing FAIR Assessment Tools. *Zenodo* <https://doi.org/10.15497/rda00035>
 28. Löbe M, Matthies F, Stäubert S, Meineke FA, Winter A (2020) Problems in FAIRifying Medical Datasets. *Stud Health Technol Inform* 16(270):392–396. <https://doi.org/10.3233/shti200189>
 29. Bonello J, Cachia E, Alfino N. (2022) AutoFAIR-A portal for automating FAIR assessments for bioinformatics resources. *Biochimica et biophysica acta. Gene regulatory mechanisms*. 1865(1):194767. <https://doi.org/10.1016/j.bbagr.2021.194767>
 30. Devaraju A, Huber R (2021) An automated solution for measuring the progress toward FAIR research data. *Patterns*. 2(11). <https://doi.org/10.1016/j.patter.2021.100370>
 31. David R, Mabile L, Yahia M et al (2018) Comment opérationnaliser et évaluer la prise en compte du concept 'FAIR' dans le partage des données: Vers une grille simplifiée d'évaluation du respect des critères FAIR <https://doi.org/10.5281/zenodo.2551500>
 32. Amdouni E, Bouazzouni S, Jonquet C (2022) O'FAIR: Ontology FAIRness Evaluator in the Agro-Portal semantic resource repository. *European Semantic Web Conference*. Springer, Cham, S 89–94 https://doi.org/10.1007/978-3-031-11609-4_17
 33. Garijo D, Corcho O, Poveda-Villalón M. (2021) FOOPS!: An Ontology Pitfall Scanner for the FAIR principles. *ISWC (Posters/Demos/Industry)*. <https://foops.linkeddata.es/about.html> Zugegriffen: 26. Februar 2024
 34. Huber R, Devaraju A (2021) F-UJI: an automated tool for the assessment and improvement of the FAIRness of research data. *EGU General Assembly Conference Abstracts* (pp. EGU21-15922). <https://www.f-ujl.net/index.php?action=test> Zugegriffen am 26. Februar 2024
 35. Cox SJ, Gonzalez-Beltran AN, Magagna B, Marinescu MC (2021) Ten simple rules for making a vocabulary FAIR. *Plos Comput Biol* 17(6):e1009041. <https://doi.org/10.1371/journal.pcbi.1009041>
 36. Erdmann C, Simons N, Otsuji R et al (2019) Top 10 FAIR Data Software. *Things*. <https://doi.org/10.5281/zenodo.2555498>
 37. Clarke DJ, Wang L, Jones A, et al. (2019) FAIRshake: toolkit to evaluate the FAIRness of research digital resources. *Cell systems*. 9(5):417–21. <https://fairshake.cloud/> Zugegriffen: 26. Februar 2024
 38. Krans NA, Ammar A, Nymark P, Willighagen EL, Bakker MI, Quik JT. (2022) FAIR assessment tools: evaluating use and performance. *Nanolmpact*. 27:100402. <https://fairaware.dans.knaw.nl/> Zugegriffen: 26. Februar 2024
 39. Inau ET, Sack J, Waltemath D, Zeleke AA. (2023) Initiatives, Concepts, and Implementation Practices of the Findable, Accessible, Interoperable, and Reusable Data Principles in Health Data Stewardship: Scoping Review. *Journal of Medical Internet Research*. 25: e45013. <https://www.jmir.org/2023/1/e45013>
 40. Sinaci AA, Núñez-Benjumea FJ, Gencturk M et al. (2020) From raw data to FAIR data: the FAIRification workflow for health research. *Methods of information in medicine*. 59(S01):e21–32. <https://doi.org/10.1055/s-0040-1713684>
 41. Welter D, Juty N, Rocca-Serra P et al (2023) FAIR in action—a flexible framework to guide FAIRification. *Sci Data* 10(1):291. <https://doi.org/10.1038/s41597-023-02167-2>
 42. Martínez-García A, Alvarez-Romero C, Román-Villarán E, Bernabeu-Wittel M, Parra-Caldéron CL (2023) FAIR principles to improve the impact on health research management outcomes. *Heliyon*. 9(5). <https://doi.org/10.1016%2Fj.heliyon.2023.e15733>
 43. Jansen C, Beier M, Witt M, Frey S, Krefting D (2017) Towards reproducible research in a biomedical collaboration platform following the FAIR guiding principles. *Companion Proceedings of the 10th International Conference on Utility and Cloud Computing*. <https://doi.org/10.1145/3147234.3148104>
 44. Jansen C, Annuschein J, Schilling B et al (2020) Curious Containers: A framework for computational reproducibility in life sciences with support for Deep Learning applications. *Future Gener Comput Syst* 112:209–227. <https://doi.org/10.1016/j.future.2020.05.007>
 45. Schindler D, Bensmann F, Dietze S, Krüger F (2022) The role of software in science: a knowledge graph-based analysis of software mentions in PubMed Central. *PeerJ Comput Sci* 8:e835. <https://doi.org/10.7717/peerj-cs.835>
 46. Klötgen M, Lauff F, Stäubert S, Meister S, Ammon D (2022) A Trusted Data Sharing Environment based on FAIR Principles and Distributed Process Execution. *HEALTHINF*, S 583–589 <https://doi.org/10.5220/0010869800003123>
 47. Hubbard T, Reilly G, Varma S, Seymour D. (2020). Trusted Research Environments (TRE) Green Paper (2.0.0). *Zenodo*. <https://doi.org/10.5281/zenodo.4594704> Zugegriffen: 8. Februar 2024
 48. Touré V, Krauss P, Gnodtke K et al (2023) FAIRification of health-related data using semantic web technologies in the Swiss Personalized Health

- Network. Sci Data 10(1):127. <https://doi.org/10.1038/s41597-023-02028-y>
49. Frexia F, Mascia C, Lianas L et al (2021) openEHR is FAIR-enabling by design. Public Health and Informatics. IOS Press, <https://ebooks.iospress.nl/doi, S113-117> <https://doi.org/10.3233/SHTI210131>

Hinweis des Verlags. Der Verlag bleibt in Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutsadressen neutral.